DOCUMENT RESUME

ED 236 213 TM 830 728

AUTHOR Wilkins, S. A.
TITLE Maximizing Student Achievement through Testing.
INSTITUTION Bossier Parish Community Coll., Bossier City, LA.
PUB DATE Mar 81
NOTE 13p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Educational Testing; Student Attitudes; *Test Use; Two Year Colleges
IDENTIFIERS *Testing Frequency

ABSTRACT

The present research attempted to determine if tests can serve as teaching devices by comparing student achievement in classes where several examinations were administered. Students in ten classes of introductory psychology served as subjects in the study. Three classes received weekly, teacher-made quizzes of 15 multiple-choice items; four classes received no examinations; three classes received three tests. The findings provide a basis for the more frequent use of examinations in teaching psychology. Examinations, or more precisely, the study they force, enables students to review and recapitulate material between classes. Although students in the weekly-tested groups did not particularly like quizzes every week, they were glad at the end of the semester that they had been tested so frequently. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236213

TM

MAXIMIZING STUDENT ACHIEVEMENT THROUGH TESTING

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. A. Wilkins

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BY

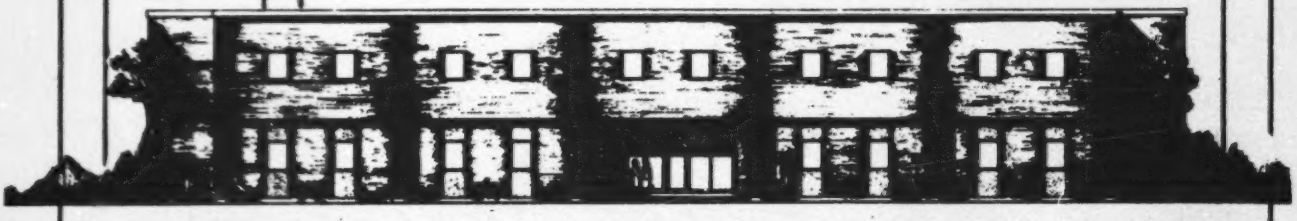
DR. S. A. WILKINS, HEAD

DEPARTMENT OF SOCIAL AND BEHAVIORAL SCIENCES

BOSSIER COMMUNITY COLLEGE

BOSSIER CITY, LOUISIANA

MARCH, 1981



bossier parish community college

■ 2719 Airline Drive North ■ Bossier City, Louisiana 71111 ■

TM830728

MAXIMIZING STUDENT ACHIEVEMENT THROUGH TESTING

Each semester college instructors spend a large proportion of time in testing students, in their classes. Throughout a student's career testing is considered to be an important, though time consuming, activity that absorbs the effort and attention of the student and professor. A fundamental question is raised: "What are the motives of the teacher in giving a test?" (Gremillion and Adams, 1977). In offering an answer to this question Phillips (1968) suggests that tests can be used to determine how much students have learned, they can be used to assign grades, and they are helpful in determining needed adjustments in instruction. In addition, tests can provide feedback, improve programs and curricula, be used in placement and selection decisions, in diagnosis, and in remediation (Sax, 1974).

While research studies give evidence supporting the use of testing in accomplishing the above, there is perhaps another reason for testing which has been little emphasized, but may be just as important. It is that ". . . evaluation, properly used, can aid learning . . ." (Tenbrink, 1974). The use of testing procedures as teaching devices has been investigated by several researchers: Phillips (1968) has reported that teacher-made tests are the most frequently used devices for gathering data about pupils. Stanley and Hopkins (1972) have stated that students consistently report greater study and learning with periodic testing, and pursue mastery of objectives more diligently if they expect to be evaluated. In taking a test a student is engaging in behavior that is instructionally valuable,

apart from the evaluation function being served by the test. Scheduled examinations stimulate review, relearning and overlearning. In a study involving test questions, Von Schilling (1972), in quoting Grossier, said that "to question well is to teach well," and that "from the time of Socrates, history's greatest teachers have generally been penetrating questioners." Gremillion and Adams (1977) have stated that studying for and preparing to take a test can be a good learning experience, since "the test gives the student individual review and a means to fill in the gaps in his own learning." According to one author, all the reasons for testing feed into this one: "to enhance student learning and growth" (Tuckman, 1975).

Many instructors feel that examinations in a lecture class are a waste of time, and that the time consumed by examination preparation could be better used in lecture preparation or research. Studies generally have not considered that examinations themselves can be effective teaching methods. The present research attempted to determine if tests can serve as teaching devices by comparing student achievement in classes where several examinations were administered. Emphasis was placed on different in-class teaching and testing frequencies with concomitant comparisons made concerning several student and class characteristics. For the purposes of this study, the term "teaching-testing frequencies" refers to the amount of time used in a classroom in teaching and the amount used in testing.

Method

Students in ten classes of introductory psychology served as subjects in the study. Students registered in the various class sections of their on volition. Five of the classes were day classes and five were night

classes, and no attempt was made to place a particular student into any section. This procedure made possible a random assignment of subjects in each section and in each treatment modality.

During the three-semester study, four classes were taught without examinations (N = 74), three classes were taught using weekly examinations were administered during a semester (N = 36). The sample for this study was limited to those students who enrolled in the researcher's classes of introductory psychology. Only those students who enrolled in the course, took the pre-test, remained in the course, and took the post-test, and for whom American College Test results were available, were included in the study. At the beginning and at the end of each semester all classes were administered the Cass Sanders Psychology Test as a pre-measure and post-measure of their knowledge of psychology. This test consists of one-hundred and twenty-five multiple choice items. In addition, students were asked to fill out a brief questionnaire. This questionnaire elicited the personal information needed to conduct the research, as well as student opinions about their examination procedure.

Post-test scores, adjusted with the use of covariance analysis, were used as achievement scores for the subjects and was the dependent variable in the study. The pre-test achievement scores and American College Test scores were included as covariables to control for pre-existing differences in student aptitude and knowledge of psychology. Students were instructed to do their best and not to guess haphazardly. They were told to answer only those questions they could, but not to spend unnecessary time on questions concerning subject-matter unfamiliar to them. Answers were recorded by placing the appropriate number on the answer sheets provided. Both pre- and post-test were administered and

scored by the researcher, and no student was told that he was participating in an experiment. The students were instructed that the pre-test was an attempt by the researcher to obtain some information on where the class stood in its knowledge of psychology, and the post-test was designed to measure how much progress has been made during the semester. Further instructions to the class members at the beginning and at the end of each semester explained that the pre-test and post-test would have no bearing on their semester grades. After the research was concluded all students were given study guides to help prepare them for the "teacher-made" final examination.

The subjects were randomly assigned according to the following schedule: one class each semester the study was conducted received weekly, teacher-made quizzes of fifteen multiple choice items. The quizzes were administered at the last class meeting of the week in each section. Four classes received no examinations. One class each semester the study was conducted received three tests. All subjects were taught by the researcher. Each class was exposed to the same amount of "in-class" lecture time on the subject matter, with the following exceptions: during the times when testing was taking place in the classes receiving three tests each semester and the classes receiving weekly tests, additional teaching was taking place in the non-tested classes. All classes were given the same reading assignments from the text, and lecture topics were identical in all classes. Additional class discussion and more indepth coverage of the subject matter consumed the additional teaching time available in the non-tested classes. The teacher-made tests consisted of multiple choice items. Some were written by the researcher while others were used that were provided by the company publishing the text. The weekly examinations consisted of fifteen multiple choice items. The classes receiving three examinations each semester were

administered tests having from fifty-five to seventy multiple choice items each. After scoring by the researcher all teacher-made examinations were briefly returned to the students. An opportunity was given for asking questions about the examinations.

The following student characteristics were obtained from the subjects and were investigated with respect to their effects on testing frequency: gender, day-night class enrollment, age, and college hours previously earned. An answer to the following specific question was sought. Is there a significant difference in achievement test scores of students taught using three tests, no tests, and weekly tests during a semester? This question and the student characteristics were restated as five null hypotheses and tested with the analysis of covariance. The null hypotheses stated that there is no significant difference in achievement test scores between subjects in the experimental groups and the control group with respect to each of the variables under study.

Results

For the purpose of analyzing the scores, the group receiving three examinations each semester was labeled the control group (C). This treatment modality was chosen as the control group because of the belief that it approximated the evaluation procedure used in a majority of college lecture-type courses similar to introductory psychology.

The group receiving weekly examinations was labeled the first experimental group (E_1), and the group receiving no examinations was labeled the second experimental group(E_2). An analysis of the adjusted post-test scores of the subjects in the experimental and control groups revealed that those subjects receiving tests, whether weekly or three per semester, had a

significantly higher adjusted mean score on the post-test than those receiving no examinations ($p < .05$). These data are presented in Table 1.

Table 1

ADJUSTED MEAN POST-TEST SCORES FOR THE

CONTROL AND EXPERIMENTAL GROUPS

GROUP	N	ADJUSTED MEAN SCORE POST-TEST	POSITIVE DIFFERENCES BETWEEN GROUPS
C	36	46.03	$C - E_2 = 3.96$*
E_1	51	45.11	$E_1 - E_2 = 3.04$*
E_2	74	42.07	

* Significant at the .05 level.

ANALYSIS OF COVARIANCE

The data derived from the analysis of covariance for the dependent variable, testing frequency, are presented in Table 2. An analysis of the socres for the total group revealed that the main effect of testing frequency was significant beyond the .05 level. Females had a significantly higher socre than males beyond the .01 level while the only significant interactions was between testing frequency and gender. All other interactions of testing frequency with the remaining variables under study failed to meet the test for significance at the .05 level of confidence. Further examination of the data revealed that the American College test scores and the Cass-Sanders Psychology Test scores were significant predictors of achievement ($p = .0001$). Although the analysis of covariance revealed a statistically significant difference between testing

Table 2

SUMMARY OF ANALYSIS OF COVARIANCE POST-TEST ACHIEVEMENT SCORES, EXPERIMENTAL AND CONTROL GROUPS

SOURCE OF VARIATION	DEGREES OF FREEDOM	MEAN SQUARES	F VALUE	PROBABILITY OF F
Testing Frequency	2	178.24	3.14	.0469
Gender	1	1127.61	19.83	.0001
Day-Night Classes	1	6.95	.12	.7271
Hours Earned	2	18.38	.32	.7244
Age	2	4.75	.08	.9199
Testing Frequency Gender Interaction	2	248.53	4.37	.0146
Testing Frequency Day-Night Class Interaction	2	85.62	1.51	.2258
Testing Frequency-Hours Earned Interaction	4	38.56	.68	.6083
Testing Frequency-Age Interaction	4	112.37	1.98	.1021
American College Test	1	3967.80	69.79	.0001
Pre-Test	1	2538.94	44.66	.0001

frequencies and a statistically significant difference between the interaction of testing frequency and gender, it was left to further analysis to show which of the groups differed significantly from the others.

THE DUNCAN'S TEST

For the purpose of determining which groups had significantly different scores, the Duncan's Multiple Range Test was chosen. It can be used to identify which groups are contributing to any overall differences if a significant F value is obtained from an analysis of covariance (Senter, 1969).

The use of the Duncan's Multiple Range Test in determining which group of subjects had the highest scores revealed that those subjects who were tested, weekly or three times per semester, had significantly higher scores that those who received no examinations ($p < .05$). There were no significant differences between those subjects tested three times per semester and those tested weekly. These data are also contained in Table 1. Since the analysis of covariance revealed no other significant interactions, further application of the Duncan's Test was unnecessary.

INTERACTION EFFECTS

A revealing finding of this study was the interaction of testing frequency with gender. After an overall significant F value was obtained, the Duncan's Test revealed that the interaction of testing frequency and gender was due to a significance difference between the females ($p < .05$). An analysis of the data showed that females scored higher than males in all groups, and females in the weekly tested group and in the group tested three times per semester scored higher than all groups of males, as well as the group of females not tested. These data are presented in table 3.

Table 3

ADJUSTED POST-TEST MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS BY GENDER

GROUP	ADJUSTED MEAN SCORES OF MALES	ADJUSTED MEAN SCORES OF FEMALES	POSITIVE DIFFERENCES BETWEEN SEXES
C	44.03 (N = 20)	48.03 (N = 16)	4.00*
E_1	39.70 (N = 17)	50.53 (N = 41)	11.83*
E_2	40.85 (N = 33)	43.27 (N = 34)	2.42

* Significant at the .05 level.

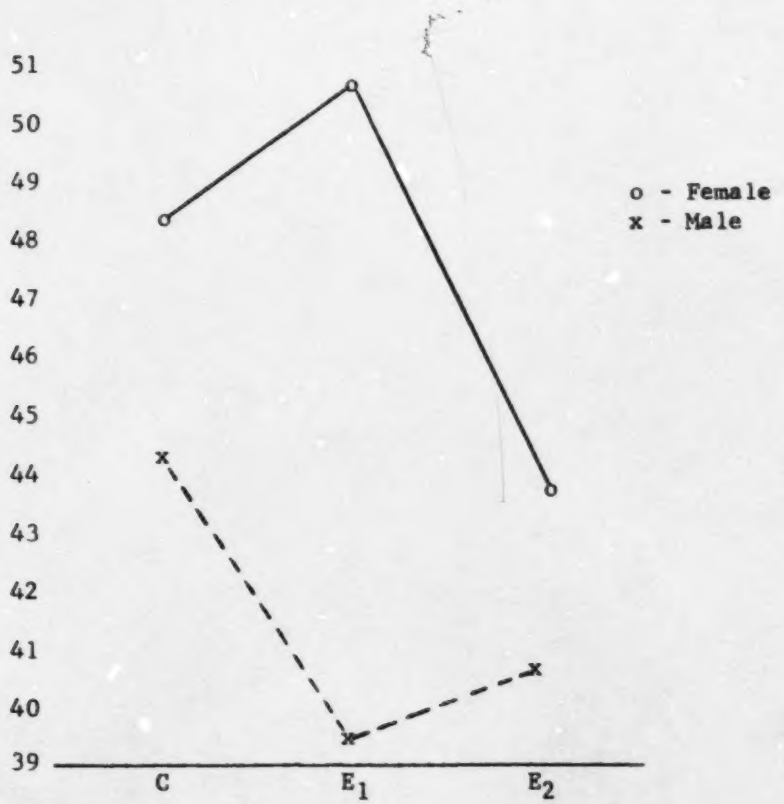


Figure 1. Interaction of Sex and Testing Frequency

Since no significant differences existed between the males at the .05 level of significance, it was concluded that the significant interaction of the main effect of testing frequency with gender was due only to the significant difference existing between the females. A graphical representation of this finding is presented in Figure 1.

Discussion.

The simple evaluation function served by examinations is the more traditional use of them. However, the amount of subject matter learned can be controlled by the number of examinations because they can force students to study. It may be an undemocratic way of coercing the student to do that which he does not want to do. The findings obtained from this investigation seem to provide a basis for the more frequent use of examinations in teaching psychology, or any similar type lecture corse. Examinations, or more precisely, the study they force, enable students to review and recapitulate material between classes. Without examinations in a lecture course, students may tend to only think about the subject matter when they come to class and listen to a lecture, and not at all between classes. By forcing repition of thoughts on the material in the mind of the student, one of the effects of an examination is to increase student learning. Quite simply stated: to learn, one must study. Examinations force students to do this. This is an implicit result of examinations that has long remained unexplored.

Perhaps a concomitant finding of this study is the student attitudes toward the corse. These attitudes were obtained both formally, from the questionnaire that was completed by the students, and informally, through student comments to the researcher and through mutual discussion of the evaluation procedure between the researcher and the students. They were not

essential to any finding of the study, but revealed some of the kinds of things that were of most concern to the students during the course of the research. The psychological effect of testing may have had more importance in influencing student achievement than the actual number of examinations. Although students in the weekly-tested groups did not particularly like quizzes every week, they were glad at the end of the semester that they had been tested so frequently. Their anxiety about the comprehensive final was less than those students who had no examinations. Informal verbal opinion sampling revealed that although students did not like the weekly tests, they preferred them to no examinations at all.

Students who were not tested until the final examination seemed to like this at the beginning of the semester, but as the final examination approached, they began to develop apprehensiveness about so much material on a single final examination. In general, most of the students receiving only the final examination did very little studying during the course of the semester, and it was these students who seemed to be the most disturbed of all the subjects in the research, regardless of testing frequency. Students who were given the more "traditional" three examinations each semester did not complain that they were given either too few or too many examinations.

Another anecdote to the experiment concerned student reaction to the standardized pre- and post-test instruments. They were somewhat discouraged by their low scores, even though it was pointed out that these scores had no effect on their grade, and even though they were encouraged not to be overly apprehensive about these scores. Although this study suggests that students profit from more frequent examinations, it is left to further research to establish what exact number of tests is optimal.